DOCUMENT RESUME

ED 267 110 TM 860 201

AUTHOR Gibble, Jacques; Masters, James
TITLE The CRT/Curriculum Articulation Project in the Penn Manor Elementary Schools. Phase One: Report of the Pilot Projects at Pequea.
PUB DATE 20 Nov 84
NOTE 39p.; Paper presented at the Annual Meeting of the Pennsylvania Association for Supervision and Curriculum Development (November 20, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; *Achievement Tests; Analysis of Covariance; Behavioral Objectives; Elementary Education; Experimental Groups; *Intervention; Language Skills; Mathematics Skills; *Pilot Projects; Pretests Posttests; *Program Evaluation; Program Implementation; Reading Skills; *Scores; Statistical Significance
IDENTIFIERS *Comprehensive Tests of Basic Skills; Penn Manor School District PA

ABSTRACT

The Penn Manor CRT/Curriculum Articulation Project was designed to study the impact of the implementation of four specific interventions on the achievement of elementary school students in mathematics and language arts. The interventions were as follows: (1) classroom teachers were given individual curriculum skill objectives for their grade assignment; (2) teachers were given individual curriculum guidelines listing essential units or lessons; (3) teachers used examples of question-answer format of achievement test in constructing classroom teacher-made tests and worksheets; and (4) test results were used diagnostically as well as descriptively at building-level as well as grade-level. Achievement was defined and measured as gains in scaled scores of the Comprehensive Test of Basic Skills. The study design compared the pre- and post-test scaled scores using an analysis of covariance statistical test. The test was chosen in an attempt to equalize any initial differences between the control and the experimental groups. The researchers found statistically significant differences favoring all experimental groups in the language arts comparisons. However, there were no statistically significant results favoring any of the experimental mathematics groups. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM

BEST COPY AVAILABLE

THE CRT/CURRICULUM ARTICULATION PROJECT IN THE PENN MANOR ELEMENTARY SCHOOLS

ED267110

PHASE ONE: REPORT OF THE PILOT PROJECT AT PEQUEA

PROJECT DIRECTOR: JACQUES GIBBLE, D. ED.

STATISTICAL ADVISOR: JAMES MASTERS, PH. D.

P.A.S.C.D. CONFERENCE

NOVEMBER 20, 1984

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

[ ] Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. Gibble

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM-860-201

ABSTRACT OF THE CRT/CURRICULUM ARTICULATION PROJECT IN THE PENN MANOR ELEMENTARY SCHOOLS

The Penn Manor CRT/Curriculum Articulation Project was designed to study the impact of the implemention of four specific interventions on the achievement of elementary school students in the areas of mathematics and language arts. Achievement was defined and measured as gains in scaled scores on McGraw-Hill's Comprehensive Test of Basic Skills (CTBS).

The four interventions were: 1) Classroom teachers were given a complete copy of the achievement test skill objectives that pertained to their grade assignment, 2) Teachers were given individual curriculum guidelines that listed certain units or lessons as "essentials." Achievement test objectives missing from the discipline guidelines at any specific level were added to the listing of "essentials" at that grade level, 3) Teachers were given examples of the question-answer format used on the achievement test so that they could use it in constructing classroom teacher-made tests and worksheets and, 4) The test results were used diagnostically as well as descriptively at building-level as well as grade-level. Formal review and goal-setting sessions were held in the fall with the building faculty on both a building- and grade-level basis.

The study design required the use of a control and an experimental group. The experimental group (all students in one school) used all four of the interventions while the control group (formed from the universe of all other students at each grade level in the district) did not use any of the interventions. The design compared the pre- and post-test scaled scores using an analysis of covariance statistical test. That test was chosen so as to equalize, as much as possible, any initial differences between the control and the experimental groups.

The researchers found statistically significant differences favoring all experimental groups in the language arts comparisons. However, there were no statistically significant results favoring any of the experimental mathematics groups. In other words, our hypotheses about the interventions were supported for language arts but not for mathematics.

We believe that the differences found in the language arts comparisons and not in the mathematics comparisons may be due to two factors. First, mathematics skills are preceived as "more imporant" than language arts skills (when the later skills are defined as English mechanics) in the mind of most elementary teachers. Therefore, the teachers may have "intutitively" made decisions about teaching "essential" skills more readily in mathematics than in language arts. If this were true, then more teachers would have taught the "essential" mathematics skills as a matter-of-course. Second, the teachers in the experimental group were required to keep a curriculum mapping log of the objectives as the objectives were taught. This may have served to increase the teachers' preception (and actions related to that preception) of their personal accountability for teaching the language arts "essential" skills.

We beleive the results to bear importance in both curriculum development at the district level and curriculum monitoring at the building level.

Researchers: Jacques Gibble, D. Ed. and James Masters Ph. D.

TABLE OF CONTENTS

PART I PURPOSE OF THE STUDY.....1

Statement of the Problem.....1

Hypotheses.....2

PART II RESEARCH CONCERNING THE INTERVENTIONS.....3

Introduction.....3

Discussion of Intervention One.....3

Discussion of Intervention Two.....6

Discussion of Intervention Three.....7

Discussion of Intervention Four.....10

Summary.....11

PART III METHODS AND PROCEDURES.....12

Research Method.....12

Data Required.....12

Method of Obtaining Data.....13

Sampling Design.....13

Analysis of Data.....13

Decisions About the Hypotheses.....14

PART IV THE FINDINGS, ANALYSIS AND DISCUSSION OF THE DATA.....15

The Study Procedures.....15

Results.....16

Discussion and Conclusions.....20

Recommendations.....21

REFERENCES.....24

APPENDICES A TO E.....25

ACKNOWLEDGEMENTS

The following people were instrumental, at strategic points, in the successful completion of this study:

Mr. Noel Taylor, Superindentent Penn Manor SD

Mr. William Rohrer, Assistant Superintendent, Penn Manor SD

Dr. Woodrow Sites, Superintendent, Donegal SD

Mr. John Sauter, Assistant Superintendent, Donegal SD

Mrs. Louise Hardinger, Principal Pequea Elementary School, Penn Manor SD

Mrs. Darla Schorr, Teacher Grandview Elementary School, Donegal SD

Teachers, Pequea Elementary School, Penn Manor SD

PART I

PURPOSE OF THE STUDY

Increasingly, standardized test scores are being used for a host of instructional and policy decisions from classifying students for remediation to providing evidence on program effectiveness. Although they are no longer just concerned with issues related to classroom practice and the use of results by teachers, these uses should still remain paramount for individual schools. Stated concisely, the most important use of achievement tests needs to be the improvement of student learning through instruction.

STATEMENT OF THE PROBLEM

During the 1982-83 school year the Penn Manor Schools began to look at patterns of achievement as evidenced on achievement tests. The district wanted to determine student growth in the areas of mathematics and language. A pre- and post-test study design used McGraw-Hill's Comprehensive Test of Basic Skills[1] in mathematics and language for the fifth and sixth grades. The study compared the growth (as computed in scaled score differences) from the spring of 1982 and the spring of 1983 using the Student's t-test. In that study, only three of the sixteen tests were significant.

This result was unexpected and disquieting. It prompted the director of elementary education to require the entire teaching staff to complete a curriculum mapping project for mathematics and language. That mapping revealed that 100% content-coverage was attained by only 5.9% of the teachers.

---

[1] Comprehensive Test of Basic Skills, Monterey CA: CTB/McGraw-Hill Publishing Co., Copyright 1982.

Additionally, it revealed that only 26.3% of the entire core curriculum was taught by 50% of the teachers.

These results expedited an earnest search for ways to improve instruction and to increase the amount of student-learning as evidenced by achievement tests. The district discovered that it, like many other school districts, had used achievement tests and, although the test results had been dutifully recorded and posted, little had been done with them in the areas of student diagnosis or in-depth program evaluation.

The superintendent appointed a project director to research the linkage of instruction, curriculum, and achievement testing. From this research came a proposal for aligning instruction, curriculum, and testing in the district.

HYPOTHESES

This study was designed to test several null hypotheses. These hypotheses were formulated to determine whether or not student-learning, as evidenced by achievement tests, could be significantly altered by a series of specific interventions.

The hypotheses were stated as follows:

Hypotheses One: There is no significant achievement difference in mathematics between an experimental group and a control group relative to specific interventions designed to align the instruction, curriculum, and testing in the district.

Hypotheses Two: There is no significant achievement difference in language between an experimental group and a control group relative to specific interventions designed to align the instruction, curriculum, and testingin the district.

PART II

RESEARCH CONCERNING THE INTERVENTIONS

The proposal written by the project director laid out several recommendations for curricular and instructional intervention. The interventions are described in this section.

Introduction

Achievement tests can be productively used to improve instruction. They can also be used to judge program effectiveness. When they are used to judge effectiveness they become a prime assessment of quality and control. Quality control has been called linking curriculum to instruction and then those two components to assessment.[2]

Discussion of Intervention One.

An achievement test can help to evaluate the common curricular experience that students have. Educators, however, need to identify the content and skills that they want the children to be uniformly exposed to at each grade. English calls this the first step in determining quality control, writing the true curriculum.[3] This content and these skills are not minimum competencies nor are they all that should be included in a quality instructional program. But, with identification of essential content and skills for each grade level, many of the uncertainties of what should be taught are minimized. As a recent researcher pointed out

> The number of facts and concepts that are part of a full curriculum is staggering. But they are not all judged to be of equal importance.

---

[2]Fenwick English and Betty Steffy "Diagnosing and Resolving Quality Control Problems in School Systems," ASCD Audio tape (612-20344) of ASCD Conference, 1983.

[3]Ibid.

Rather, some items are singled our for special emphasis.[4]

But how are teachers to determine what skills should be singled out as most important? How are they to resolve the differences in content messages communicated through textbooks and tests when deciding what to teach? Are they to guess? Some educators, like Shilman, feel that this is a moot question. He has postulated

> Teachers do not focus on outcomes. They do not focus on objectives. They do not focus on goals, though the goals are probably there implicitly. Teachers focus on activities and content. Their attention is directed to questions of what will we do and what will we cover.[5]

Most teachers rely on the textbook to provide adequate skill coverage and are unaware that the skills that are tested on an achievement test may be different. Thus, while we see teachers teaching, it may be that only a small fraction of what they teach is in any way linked to the tested curriculum. The Instructional Dimensions Study looked at the effects of instruction on achievement. It claimed that it is important to expose students to the content of the test before the test is administered. In fact, achievement gains were reported to be more highly correlated with the amount of test content taught than with any other classroom variable. Yet it reported the overlap of content taught to content tested ranged from 4-95%. The researchers reported that at least 60% of the content tested by the test has to be taught to mastery just so the students score at grade level. Cooley and Leinhardt conclude:

> Instructional techniques, both curricular and interpersonal, donot compensate for missing content. Children perform better on a test if they have been exposed to both the form of the test items and the content

---

[4]Robert Linn, "Testing and Instruction: Links and Distinctions," Journal of Educational Measurement. Summer, 1983, p. 179.

[5]L. S. Shilman. Test Design: A View From Practice. U.S. Educational Resources Information Center, ERIC Document ED 173 391, June, 1978, p. 10.

covered by the test. This point may seem trivial but it is in fact quite important.

> Both what is taught and what is tested represent samplings from domains. Curriculum and test can clearly emphasize different content. The results of IDS show that students are most likely to answer correctly if they have been taught the specific material covered by a test and if they have been frequently exposed to the test format.[6]

While we might agree that the number of skills may be overwhelming to some teachers' planning, we would also agree that a central core of sensible, well-chosen objectives can help the teachers focus on goals. Few curriculum guides and teaching and testing cannot be said to be well "aligned" or "linked" or "coordinated" or "articulated."

Curriculum, instruction and assessment do not significantly overlap. Because all objectives in the curriculum are all equally important, a teacher must really guess about which ones should be taught.[7] To resolve this conundrum Haetrel and Calfee recommended

> A better strategy [than guessing]...would be to provide the teacher with the actual design specifications for the test battery at the level of item domain descriptions. The teacher would then have no need to guess, and could even use items from the various domains for planning the instructional program, taking care to stay away from those items on the final exam.[8]

This step would make more probable the achievement of the second and third steps of English's quality control monitoring, teaching to the written cur-

---

[6]M. E. Brady, et. al., Instructional Dimensions Study. Washington, D.C.: Kirschner Associates, Inc., 1977.

[7]William Brock and Roger Scott. "Improving Instruction in Essential Skills," What's Noteworthy on School Improvement and Technology. Winter, 1982, p. 85.

[8]E. Haertel and R. Calfee. "School Achievement: Thinking About What to Test," Journal of Educational Measurement. Summer, 1983: p. 126.

BEST COPY AVAILABLE

riculum, and testing the content taught.[9] The concern here is that the district administration, perhaps in conjunction with teachers, make this judgment.

Intervention One: Classroom teachers will be given a complete copy of the achievement test skill objectives that pertain to their grade assignment.

Discussion of Intervention Two

The problem of giving teachers the listing of the test skills as well as the listing of the skills and content to be taught at each grade is not difficult. But, it is extremely important that when this is done the two sets of objectives are cross-referenced to make sure that all the children are exposed to the objectives measured on achievement tests.

Thematchof "classroom" content taught and tested and content tested on achievement tests is not great. Numerous research pieces have demonstrated that the coordination between tests and curriculum development is often small.[10] A recent study comparing one district's curriculum objectives in both language and mathematics with the state assessment and various standardized tests illustrates this irrationality. Some tests included only 25% of the district curriculum, and, in some, almost 50% of the test was not covered in the curriculum.[11]

A study conducted by the Institute for Research on Teaching (IRT) at Michigan State found that

---

[9]English, op. cit.

[10]McREL. "How Well Do Tests, Textbooks Match?" What's Noteworthy on School Improvement. Summer, 1981, p. 45.

[11]E. Baker and J. Herman. "Task Structure Design: Beyond Linkage," Journal of Educational Measurement. Summer, 1983. p. 126.

BEST COPY AVAILABLE

> ...there is substantial variation across achievement tests in content coverage... any given textbook examined covered only slightly more that half of the [content covered in achievement tests]...[12]

The study added "...even in the best matched textbooks and tests, less than 75% of the specific topics...on the exam are...in the book..."[13] Brock summarized the dilemma well when he added

> The principle is this: we should teach students what we expect them to learn, and we should only test students on what they have had an opportunity to learn...classroom instruction covers all of the instructional objectives. The assessment samples the student accomplishments and does not measure anything other than the stated objectives of the program.[14]

The consideration of content and construct validity is necessary here if we are going to link performance on the test to instruction in the classrooms. "Choosing content that matches the outcome measure," wrote Berliner "...is a critical and often overlooked instructional decision. Congruence between what is taught and what is tested must be high."[15] Linn gave a concrete example of this when he wrote about a district that gave two different achievement tests to their primary students in two consecutive years. The mean gain was 12 points. Under analysis it was discovered that the test in the second year contained 55 percent of the vocabulary developed at that grade level while the test in the first year contained only 19 percent.[16]

---

[12]McREL, op. cit., p. 44.

[13]D. J. Freeman, et. al. The Fourth Grade Mathematics Curriculum As Inferred from Textbooks and Tests. U.S. Education Resources Information Center, ERIC Document ED 199047, July, 1980, p. 15.

[14]Brock, op. cit.

[15]David Berliner, "Viewing the Teacher as a Manager of Decision," Education Digest. February, 1982, p. 20.

[16]Linn, op. cit., p. 180.

BEST COPY AVAILABLE

Clearly, the match between test content and curriculum content matters. Floden concurred and wrote "...a change in test forms can lead to substantially different grade equivalent scores...the choice of test can greatly influence the estimate of program effects."[17] His analysis made more apparent the significance of these differences. He found that a single correct answer on the SAT Math subtest can add .2 to a students G.E. score, it it is near the middle of the norm distribution.[18]

Floden wrote "...content covered by the test [should] match the content of most concern to those using the test results."[19] Thus, some "additional" skills may need to be taught if the test and district/school interpretation of it is to have instructional validity. Airasian wrote, "Instructional validity goes one step further in referring to the match between test content and the content of the instructional actually provided in the classrooms."[20] Linn extended this concern when he reported

> The importance of a close link between the content of a test and the content of instruction was forcefully underscored by the Fifth Circuit Court of Appeals in the Debra P. v. Turlington decision. The court strongly emphasized the need to ensure that the topics on the test are actually taught. This is evident in the following statement from that decision,'We believe that the state administered a test that was...fundamentally unfair in that it **may** have covered matters not taught in the schools of the state.[21]

---

[17]R. E. Floden. Don't They All Measure the Same Thing? Consequences of Selecting Standardized Tests. U.S. Education Resources Information Center, ERIC Document ED 167 632, July, 1978, p. 7.

[18]Ibid.

[19]Ibid., p. 13.

[20]Peter Airasian and George Madaus. "Linking Testing and Instruction: Policy Issues." Journal of Educational Measurement. Summer, 1983, p. 107.

[21]Linn, op. cit., p. 185.

BEST COPY AVAILABLE

Haetrel and Calfee summarized

> And here is the crux of the matter, for the only guide we can use to establish the validity of a test is the validity of the instructional program itself. To the degree that we lack certainty about what is to be accomplished in the classroom and in the mind of the student, then the criteria for establishing test validity is undercut."[22]

Intervention Two: Teachers will be given individual discipline guidelines that list certain units or lessons as "essentials." Achievement test objectives missing from the discipline guidelines at any specific level will be added to the listing of "essentials" at that grade level.

Discussion of Intervention Three

The overlap between what is taught and what is tested is an important predictor of test performance. Recently, research has been completed that indicates that students exposed to achievement "test format" do better on those types of tests. Both the IDS and Carlburg reported that teachers should teach the skills in test-taking. Carlburg's concept of "test-wiseness" has a great effect on timed tests like achievement tests. He claimed that teaching students how to mechanically handle tests will increase their scores by 8 points at the 40th percentile, 6 points at the 50th percentile, and 5 points at the 80th percentile.[23]

Test authors build into tests difficulty factors that are not related to how much a child knows but to how well he or she can think in a testing situation. These difficulty factors, such as vertical vs. horizontal math problems, multiple choice math answers, etc.can be learned as a mechanical form. The negation of this type of mechanical difficult factor, said Linn

---

[22]Haetrel and Calfee, op. cit., p. 127.

[23]Conrad Carlburg, "The Testing Game: Minimizing the Irrelevant Factors," What's Noteworthy on School Improvement. Summer, 1981, p. 59.

BEST COPY AVAILABLE

"...is apt to yield results that are more sensitive to the instruction that is taking place..."[24] The significance of this issue is that specific material should be taught and practiced in the test's format during classroom instruction so that students become familiar with it.

Intervention Three: Teachers will be given examples of the question-answer format used on the achievement test so that they can use it in constructing classroom teacher-made tests and worksheets.

Discussion of Intervention Four

Historically, the usefulness of achievement testing has been mostly in the area of description. Often little has been done with the test results in the area of diagnosis of children's problems and even less with program evaluation. Linn disclosed that

> ...direct instructional uses of standardized tests have been rather modest. After a test is administered, the scores are often recorded, reported and forgotten until the next year when the process is repeated. Clearly, there are exceptions...[but the] present day testing programs are largely extraneous to everyday classroom teaching...[c]urrent testing procedures are not helpful to teachers or students in their day-to-day efforts to teach and learn.[25]

Then, in describing earlier research, he wrote that this resulted in "...teachers value[ing] their own judgment more highly than information from tests which were found to be little used in instructional decision-making."[26]

Teachers continue to feel that the social make-up of a class contributes greatly to test performance. Numerous data tests done by Airasian indicate that most of the variance on sub-tests within a school "...was due to instruc-

---

[24]Linn. Op. cit. p. 184.

[25]Ibid.

[26]Ibid.

BEST COPY AVAILABLE

tional variation rather than to general ability or social class variation."[27] This former type of opinion is promoted when tests are not discussed and it is definitely not what we desire.

Intervention Four: The test results will be used diagnostically as well as descriptively at building-level as well as grade-level. Formal review and goal-setting sessions will be held in the fall with the building faculty on both a building- and grade-level basis.

SUMMARY

Evaluation is a necessity in that it gives vital information regarding the progress of the student and the school itself. Edmonds wrote,

> Despite all of the limitations of standardized test, I would argue as forcefully as I can that they are--at this moment--the most realistic, accurate, and equitable basis for portraying individual pupil progress.[28]

We have come to know a great deal about achievement testing. We can have the type of achievement monitoring program we want if we make some changes. These changes, when fully implemented, will help enable us to reach our ultimate goal of the improvement of student learning through instruction.

---

[27]Airisian, Op. cit. p. 180.

[28]As quoted in Effective Schools. Vol. 1, No. 1 (December, 1983), p. 4.

BEST COPY AVAILABLE

PART III

METHODS AND PROCEDURES

The purpose of this study was to evaluate the effect of certain intervention procedures on achievement test scaled scores. This section describes the methods and procedures used in achieving this purspose.

RESEARCH METHOD

This study used a pretest-posttest control group design. Leedy recommended this type of design when the researcher desired strong internal and external validity.[29]

DATA REQUIRED

This study worked only with the mathematics and language subtests of the McGraw-Hill Comprehensive Test of Basic Skills (CTBS). It required that two types of information be collected.

First, it required the composite scaled scores from the mathematics and language subtests from the spring, 1983 CTBS test administration (to be used as pretest scores) and the composite scaled scores from the mathematics and language subtests of the spring, 1984 CTBS test administration (to be used as posttest scores). This data would be used for the statistical analysis.

Second, it required that data be gathered to ascertain the percentage of "essential" objectives that had been taught. This information would be used descriptively to discuss the relationship of objectives covered to the raw gains on the scaled scores.

---

[29]Paul D. Leedy, Practical Research: Planning and Design. (New York: Macmillian, 1974), p. 151.

BEST COPY AVAILABLE

METHOD OF OBTAINING DATA

Data for the pretest-postest analysis was obtained from CTBS computer printouts for both the spring, 1983 and spring, 1984 test administrations.

Data used for the determination of objective coverage was gathered using the "Curriculum Mapping Log" form found in Appendix A.

SAMPLING DESIGN

The experimental group for mathematics was designated as all class members of grades 2 through 6 at the Pequea Elementary School in the Penn Manor School District. The experimental group for lanugage arts was designated as all class members of grades 3 through 6 at the Pequea Elementary School of the Penn Manor School District. The control groups were established for each grade and subject by using a universe of all students in the district at that grade level and then randomly selecting a number of students equal to the number in the experimental group.

ANALYSIS OF DATA

The main body of the test required that scaled scores obtained from the mathematics and language sub-tests of the CTBS be examined using the analysis of covariance statistical test. This test allows the posttest means to be compared using the pretest means as a covariate. If the variation cannot reasonably be attributed to sampling error, the researcher can conclude the the treatments applied had an effect. This test is recommended in those instances when the researcher has reason to believe that the experimental and control groups are not statisically equal.

Each grade level was compared for mathematics and then for language using the analysis of covariance test. A multiple factorial comparison test was not done because there was adequate justification for doing each grade level

separately (i.e. each grade used a different battery of the CTBS).

The second part of the study required teachers to log each objective as it was taught to the students. From these logs, the percentages of total objective coverages were determined for each grade level. These percentages were then descriptively correlated with raw gains on the scaled scores.

DECISIONS ABOUT THE HYPOTHESES

For the analysis of covariance test an obtained value for the F statistic which had the probability of occurrence of less than .05 was considered significant. For the comparison of the gain in scaled score and percentage of "essential" objectives covered during the year, a correlation would have served well. However, the number of cases made a statistical correlation impractical. We felt such a correlation would be to unstable. Therefore, it was decided to handle this material descriptively.

PART IV

THE FINDINGS, ANALYSIS AND DISCUSSION OF THE DATA

This section presents the data from the study. To describe the study succinctly but comprehensively this section is divided into four major sections: The Study Procedures, The Data Findings, Analysis and Discussion,

THE STUDY PROCEDURES

The study itself was designed to evaluate the effectiveness, in terms of scaled scores gains on the CTBS achievement test, of four interventions that were described in Part II. The method for implementing the interventions is described in this first section.

During the summer of 1983 the project director spent several weeks isolating the CTBS skill objectives for the mathematics and language subtests and articulating them with the textbooks used as the elementary curriculum (the planned courses were only begun in fall of 1983). This articulation gave each teacher a complete list of the CTBS ojectives tested at their grade level, cross-referenced the objectives with the curriculum so that skills tested but not taught at a particular grade level were listed as required for teaching at that level, and gave teachers a sample of the types and formats of the test questions used at their grade level.

The resulting document was prefaced with a cover sheet (see Appendix B) for the mathematics section and another for the language section that listed certain lessons as "essentials" with others considered "extensions" to be taught only after the "essentials were taught." The cover sheet stressed the requirement that the teachers should teach one to three lessons on each "essential" skill that was not already covered in the curriculum guidelines.

The document (a portion of which is in Appendix C) was printed for distribution to teachers during a work session in August before the beginning

of school. At that session the project director described the project's goals and each teacher's responsibilities.

The curriculum handbook was examined and explained in detail and the results of the spring, 1983 testing were formally discussed at both the building and individual grade level. The teachers were introduced to the objective logging system that would be used in the study and were told that quarterly meetings would be held to answer questions related to the study and the teachers' responsibilities in it.

This initial meeting with all of the teachers in the experimental group was followed with three quarterly meetings where the project director answered questions and encouraged the teachers to follow the curriculum and intervention guidelines. At these meetings the project director collected the objective logs from each teacher. The last meeting was held one week prior to the beginning of the achievement testing.

The scaled scores for the students in both the experimental and control groups were isolated and gathered in June, 1984. In six cases, students initally included in the experimental group were removed because they had moved from the district.

RESULTS

The results for the language arts subtests are summarized in Table 1 below. Shown in the table are pre- and post-test means, standard deviations, the adjusted post-test means compared in the analysis of covariance, and the F values for those comparisons. (Analysis of covariance source tables may be found in Appendix D.)

Table 1

Mean Scores and Analysis of Covariance F Values for Control and Experimental Groups in Language Arts

Grade	Group	n	Pretest Mean	Pretest S.D.	Posttest Mean	Posttest S.D.	Adj. Mean	F Value
3	Con.	21	623.74	55.85	664.38	41.33	664.38	2.94*
	Exp.	21	595.23	50.19	645.38	52.38	645.29	
4	Con.	15	662.20	55.29	704.88	36.90	704.15	2.78*
	Exp.	15	659.88	42.76	689.20	38.99	689.91	
5	Con.	24	720.00	37.35	754.54	62.77	734.69	6.49*
	Exp.	24	679.88	41.12	687.63	34.44	707.47	
6	Con.	34	707.62	25.05	743.22	34.65	740.22	13.24*
	Exp.	34	699.74	31.46	716.43	27.06	719.43	

* $p < .05$

The analysis of data for language arts yeilded significant values at all levels. The values reached or exceeded the critical values and were significant at beyond the .05 level.

The results for the mathematics subtest are summarized in Table 2 below. Shown in the table are pre- and post-test means, standard deviations, the adjusted post-test means compared in the analysis of covariance, and the F-values for those comparisons. (Analysis of covariance source tables may be found in Appendix D.)

Table 2

Mean Scores and Analysis of Covariance F Values for Control and Experimental Groups in Mathematics

Grade	Group	n	Pretest Mean	Pretest S.D.	Posttest Mean	Posttest S.D.	Adj. Mean	F Value
2	Con.	25	509.96	84.05	609.64	52.93	616.35	0.21
	Exp.	25	539.56	67.98	628.52	49.39	621.81	
3	Con.	21	631.91	59.11	655.19	36.94	658.58	2.66
	Exp.	21	645.83	59.64	675.57	31.43	672.18	
4	Con.	15	677.20	45.35	692.33	24.30	689.17	0.91
	Exp.	15	664.13	39.66	691.73	27.71	691.73	
5	Con.	24	703.75	20.92	715.67	22.75	703.87	0.85
	Exp.	24	671.58	23.34	686.04	25.60	697.84	
6	Con.	35	709.14	23.14	709.14	23.08	715.54	0.19
	Exp.	35	702.80	18.30	714.37	16.49	716.69	

* $p < .05$

The analysis of data for mathematics yeilded no significant values at any levels. No F value reached or exceeded the critical values and, therefore, none were considered significant.

The scaled score gains and percentage of objectives covered data for language are summarized in Tables 3. The scores for 1983 and 1984 are for identical members of the experimental group.

Table 3

Scaled Score Gains and Percentage of Objective Coverage by Grade Level for Language

Grade Level	Average Scaled Score 1983	Average Scaled Score 1984	Average Raw Gain	Percentage of Coverage
3	623	664	41	100
4	662	694	32	92
5	725	729	4	46
6	708	735	27	92

The scaled score gains and percentage of objectives covered data for mathematics are summarized in Table 4. The scores for 1983 and 1984 are for identical members of the experimental group. As can be seen from these tables, in Language Arts, the classes which had the greatest, in general, averages gains. This did not occur in mathematics as can be seen below.

Table 4

Scaled Score Gains and Percentage of Objective Coverage by Grade Level for Mathematics

Grade	Average Scaled Score 1983	Average Scaled Score 1984	Average Raw Gain	Percentage of Coverage
2	506	609	103	78
3	631	655	24	66
4	676	692	16	93
5	703	715	12	71
6	708	717	9	93

DISCUSSION AND CONCLUSIONS

Analysis of data related to language arts yielded statistically significant F values ateach grade level. Since these values exceeded the critical value it was decided that hypothesis number two would be rejected and the alternative hypothesis would be accepted. The alternative hypothesis was that "There is a significant difference in language achievement between an experimental group and a control group relative to specific interventions designed to align the instruction, curriculum and testing in the district."

The analysisof data related to mathematics yielded no statistically significant value at any grade level. Since none of these values reached or exceeded the critical value it was decided that hypothesis number one would be retained.

The data gathered on percentage of objectives covered was not handled in a statistical manner. A descriptive analysis of the language data would seem to indicate that the higher the percentage of the objectives taught the higher the average scaled score gain. In all four cases, the higher the percentage of objectives taught, the higher the scaled score gain.

A descriptive analysis of the mathematics data appear to show that there is little connection between the percentage of objective coverage and gain in averages on the scaled score. The largest gains are at the lower grades but the highest percentage of time spent on teaching the objectives are at the upper grades. The "implied" correlation is that of negative direction.

The study results indicate that the use of the four interventions did have a positive, and significant result on the achievement of the students in the area of Langauge Arts. Our analysis is that the increased emphasis on specific "essential" objectives, heightened by the fall school- and grade-level meetings as well as the printing of a specific curriculum objective guide, helped teachers focus their efforts more accurately on achieving the

project's curriculum objectives. We also feel the increased accountability required as teachers completed and turned in the "Curriculum Mapping Log" raised their level of concern for working with the project's goal objectives. We believe that the use of the interventions had a synergistic effect on the test results. In other words, using only one or two of the interventions may not have achieved the level that was achieved when the interventions were used in combination. There is, of course, no way to determine the amount of increased achievement that could be attributed to any one of the interventions singly.

The teachers in the experimental group showed initial and continual anxiety over the use of the required "logging" system for tracking the objectives that had been taught. We believe that this increased anxiety and its attendant high awareness may have caused the teachers to increase their efforts to cover the "essential" objectives to a higher level of mastery than they might have otherwise.

Such a finding might also partially explain the lack of significance in the mathematics test areas. Traditionally, mathematics has experienced a higher prestiage level than language arts (when interpreted as English mechanics skills). Teachers may attend more closely to coverage of what they "intuitively" consider the "essentials" in mathematics than in language. The result of this would be that the teachers might have "intuitively" taught more of the essential objectives to a high level of mastery simply as a matter-of-course. In other words, the language scores may have increased because: 1) the teachers <u>now</u> felt that area to be more important and, 2) their personal accountability, in the form of the logging system, was being more closely scrutinized by the project staff.

While the achievement in the experimental group's mathematics scores was not increased significantly over those of the control group, it should be

noted that the experimental groups' achievement was not decreased by use of the interventions. Once again, this may be the result of a higher level of teacherperceived accountability and feelings of significance of this subject.

To test out the possibility that teachers in the mathematics control groups may "intuitively" have taught the "essential" objectives as a matter-of-course and thus achieved results similar to the experimental groups more (supposedly) focused teachingefforts, the pre- and post-test scores were examined independently using a Student's t-test. The t-tests were also completed for the language arts groups for descriptively comparative purposes. (These t-test tables are in appendix E).

In mathematics, all but one control group and all of the experimental groups achieved statistically significant scores. In language arts only one control group, but all of the experimental groups achieved statistically significant results. These results would seem to confirm our analysis that the teachers in mathematics "intuitively" teach the "essential" objectives. In language, perhaps because of its rather non-sequential nature, the control group teachers apparently did use the same criteria for their content/teaching decision-making.

The descriptive analysis of objectives covered and scaled score gain visually "appears" to correlate with the increased achievement in language. That is, the higher the percentage of objectives, the higher the scaled score gain. The higher the scaled score gain, the higher the statistical significance of the analysis of covariance test.

RECOMMENDATIONS

We feel rather confident that the use of the four intervention strategies examined in this study would be worthy of consideration in other curriculum implementation efforts. The results lead us to believe that these efforts may

give more pronounced results in subject areas other than mathematics. Mathematics, both because of its high profile in elementary curriculum and its critically sequential nature, may benefit less than areas like Language Arts or Social Studies, which are less critically sequential. Although, it should be noted that the results for the experimental group were at least as impressive in mathematics as in Language Arts.

REFERENCES

Airasian, Peter and George Madaus. "Linking Testing Instruction: Policy Issues." Journal of Educational Measurement. Summer, 1983, p. 107.

Baker, E. and J. Herman. "Task Structure Design: Beyond Linkage," Journal of Educational Measurement. Summer, 1983.

Berliner, David. "Viewing the Teacher as a Manager of Decision," Education Digest. February, 1982.

Brady, E. , Instructional Dimensions Study. Washington, D.C.: Kirschner Associates, Inc., 1977.

Brock, William and Roger Scott. "Improving Instruction in Essential Skills," What's Noteworthy on School Improvement and Technology. Winter, 1982.

Carlburg, C. "The Testing Game: Minimizing the Irrelevant Factors," What's Noteworthy on School Improvement. Summer, 1981, p. 59.

Effective Schools. Vol. 1, No. 1 (December, 1983).

English, Fenwick and Betty Steffy., "Diagnosing and Resolving Qaulity Control Problems in Public Schools."" ASCD Audio tape (612-20344) of ASCD Conference, 1983.

Floden, R. E. Don't They All Measure the Same Thing? Consequences of Selecting Standardized Tests. U.S. Education Resources Information Center, ERIC Document ED 167 632, July, 1978, p. 7.

Freeman, D. J. et. al. The Fourth Grade Mathematics Curriculum As Inferred from Textbooks and Tests. U.S. Education Resources Information Center, ERIC Document ED 199047, July, 1980.

Haertel, E. and R. Calfee. "School Achievement: Thinking About What to Test," Journal of Educational Measurement. Summer, 1983.

Linn, Robert. "Testing and Instruction: Links and Distinctions," Journal of Educational Measurement. Summer, 1983.

McREL. "How Well Do Tests, Textbooks Match?" What's Noteworthy on School Improvement. Summer, 1981.

Shilman, S. Test Design: A View From Practice. U.S. Educational Resources Information Center, ERIC Document ED 173 391, June, 1978.

APPENDIX A

Curriculum Mapping Log for _____ for Grade_____

Obj. No. PM/CTBS	Date Taught	Ex. Prt. 80%	Comments

APPENDIX B

LANGUAGE OBJECTIVES AND CORRELATIONS

CTBS Objectives Tested at this grade-level: 5

Scott, Foresman Objectives included at this grade-level: 21

Penn Manor Primary Objectives: 11

Penn Manor Secondary Objectives: 10

Dear Teacher of Grade Six,

The project staff has correlated the Scott Foresman Language Program with the CTBS Achievement Test for your grade-level. The correlation resulted in the identification of Primary Objectives listed here and Secondary Objectives. The descriptions of the Primary Objectives can be found on pages T35-T38 of your teacher's manual.

Primary Objectives:

4*, 5*, 6*, 7*, 8*, 10*, 11*, 12*, 14, 15, 16, 17*, 18*, 19*, 20*, 22*, 24*, 26*, 30, 31, 32*, 33*, 35*, 36*, 39*

The Primary Objectives are considered crucial and all should be taught to mastery level (80%-100% for each child) before going on to the Secondary Objectives.

The format of the sample problem(s) is the one that you should try to use in preparing worksheets and tests. The Unit Tests and worksheets in the program are acceptable as they are but it will probably be necesssary to create additional ones.

To help you in the creation of a yearly plan we have taken each text objective that is Primary and indexed the text pages and a sample problem in the white pages that follow. The colored pages include the indexed CTBS objectives and correlated textbook objectives.

You may want to use a Mastery Learning strategy or a Student Team Learning procedure to help insure higher levels of student success. The project staff is available to consult with you and help you plan for a successful year.

Sincerely,

CTBS Correlation Project
August, 1983

APPENDIX C

Grade 6 CTBS 39

Objective: The student will identify the correct use of a capital letter for the pronoun I, proper nouns, and proper adjectives. (proper name or initials, proper a;djectives, nationality, geographical name, historical event.)

Text Reference:

Language Related Objective: None

Related pages: 80-81, 164-165

Sample Problem:

Decide which part of the sentence, if any has a word that needs a capital letter.

John smith has decided to become a lawyer for a bank. None

A B C D E F

++

Grade 6 CTBS 41

Objective: The student will identify the correct use of a period, question mark, exclamation point, ro a comma in a sentence. (comma-series, direct address, quotes, compound sentence, sentence interrupter, introductory word or phrase, introductory clause)

Text Reference:

Language Related Objective(s): 7

#7 pp. 21-22, 24-25-f

Related pages: 46-47

Sample Problem:

Decide which punctuation mark, if any, is needeD in the sentence.

The leader announced "We will take a short break."

A . B ; C , D : E None

Table 4

Analysis of Covariance Data for
Language Arts at Grade Six

Source	S.S.'	df	M.S.'	F
Treatment (Exp. vs. Con.)	7404.51	1	7404.51	13.24*
Within	37460.40	67	559.11	
Total	44864.91	68		

*p<.05

Table 5

Analysis of Covariance Data for
Mathematics at Grade Two

Source	S.S.'	df	M.S.'	F
Treatment (Exp. vs. Con.)	356.84	1	356.82	0.21
Within	78324.56	47	1666.48	
Total	78681.40	49		

Table 6

Analysis of Covariance Data for
Mathematics at Grade Three

Source	S.S.'	df	M.S.'	F
Treatment (Exp. vs. Con.)	1894.93	1	1894.93	2.69
Within	27412.92	39	702.90	
Total	29307.85	40		

Table 7

Analysis of Covariance Data for
Mathematics at Grade Four

Source	S.S.'	df	M.S.'	F
Treatment (Exp. vs. Con.)	239.62	1	239.62	0.91
Within	7109.61	27	263.32	
Total	7348.69	28		

Table 8

Analysis of Covariance Data for
Mathematics at Grade Five

Source	S.S.'	df	M.S.'	F
Treatment (Exp. vs. Con.)	281.08	1	281.08	0.85
Within	14816.78	45	329.08	
Total	15097.86	46		

Table 9

Analysis of Covariance Data for
Mathematics at Grade Six

Source	S.S.'	df	M.S.'	F
Treatment (Exp. vs. Con.)	22.29	1	22.29	0.19
Within	7818.54	67	116.69	
Total	7870.83	68		

APPENDIX E

Table 1

T-Test Summary for all Control
and Experimental Groups by Discipline

Language

Grade	Group	T-Test Score	df	Significance**
3	Con.	-1.48095	40	No
	Exp.	-2.61713	40	Yes
4	Con.	-0.96318	28	No
	Exp.	-2.48584	28	Yes
5	Con.	-0.707897	46	No
	Exp.	-2.31682	46	Yes
6	Con.	-2.30384	70	Yes
	Exp.	-5.01726	70	Yes

** p < .05

Mathematics

Grade	Group	T-Test Score	df	Significance**
2	Con.	-5.16883	48	Yes
	Exp.	-5.29320	48	Yes
3	Con.	-1.53090	48	No*
	Exp.	-1.99272	48	Yes
4	Con.	-1.13927	28	No
	Exp.	-2.20911	28	Yes
5	Con.	-1.88899	46	Yes
	Exp.	-2.04498	46	Yes
6	Con.	-1.66969	68	No*
	Exp.	-2.77914	68	Yes

** p < .05
* p < .10